

On the Influence of Geospatial Context on Mobile Microblogging Contents

Stefan Hahmann, Dirk Burghardt

Institute for Cartography, TU Dresden, Germany
stefan.hahmann@tu-dresden.de, dirk.burghardt@tu-dresden.de

Keywords: Geographic Information Retrieval, Volunteered Geographic Information, User Generated Content, Twitter, Microblogging

1. Introduction

Since the micro-blogging service Twitter has been launched in 2006, it has become very popular among a wide community of users. It is estimated that as of 2012 there are over 500 million users who generate approximately 350 million posts per day. Public posts may be queried by an API provided by Twitter. About 1% of all posts are tagged with GPS coordinates. This makes them an interesting source for the field of geographical analysis with volunteered geographic information (VGI). Figure 1 and figure 2 visualise this data in the area of Dresden.

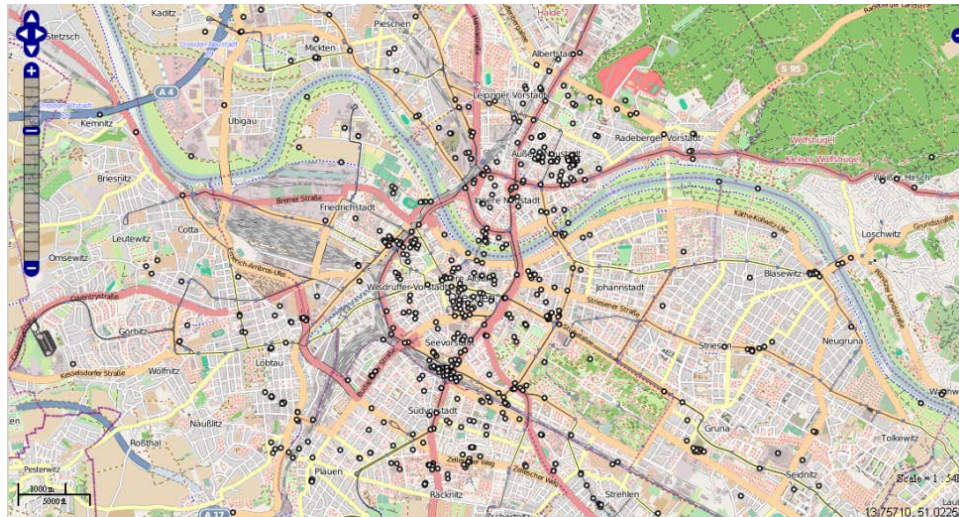


Figure 1. Distribution of recorded georeferenced posts (German language only) in the area of Dresden

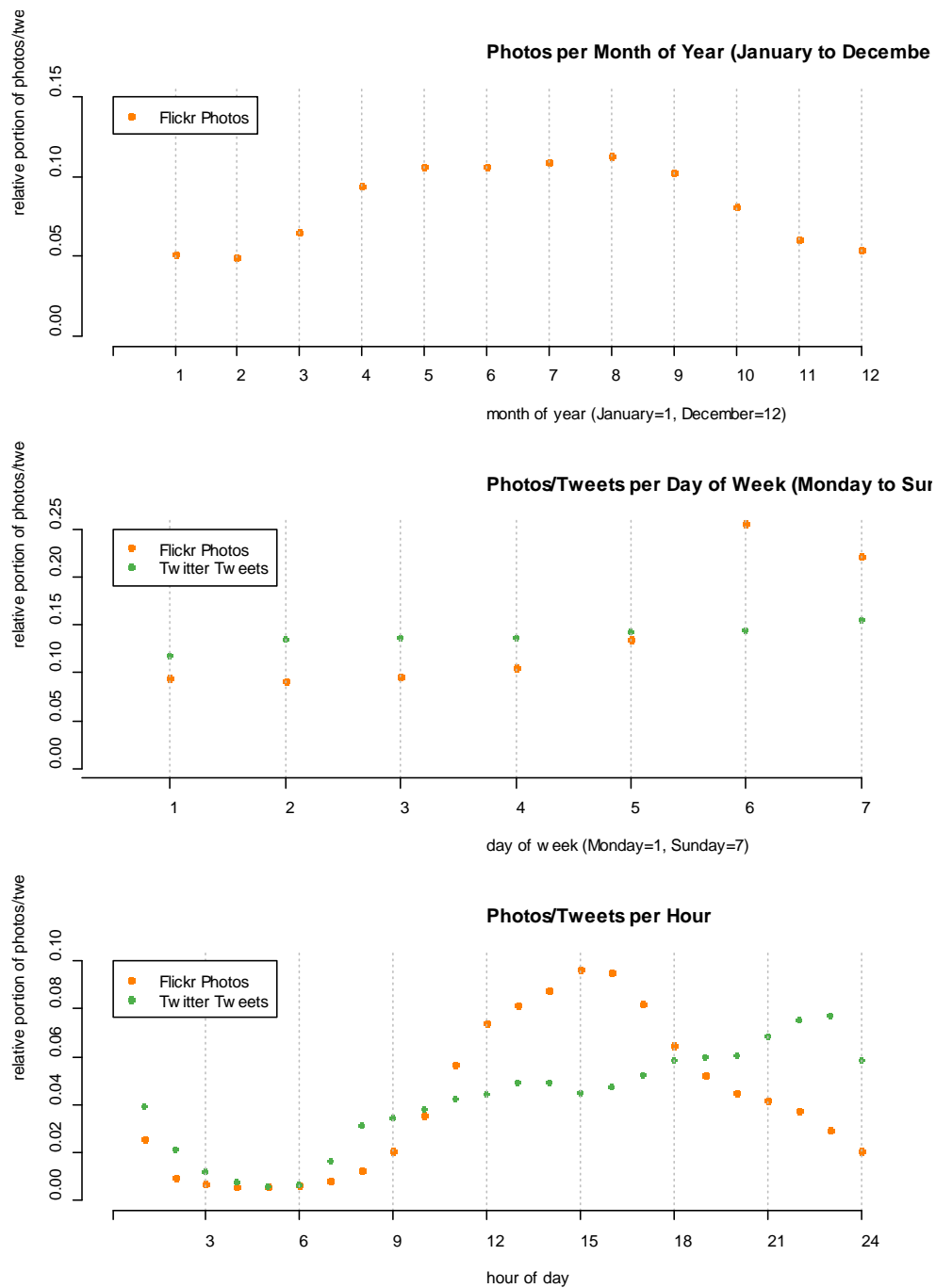


Figure 3. Temporal distribution of georeferenced Flickr photos and Twitter microblogging posts. All data have been collected within the area of Germany.

3. Correlation between Geospatial Context and Mobile Microblogging Contents

Preliminary analyses show that a certain amount of mobile microblogging posts has a relation to the place where they have been generated. For example, posts related to public transport may be found near to train stations and posts related to movies may be found near to cinemas.

Kwak et al. (2010) show that the microblogging service Twitter has properties of both, a social network and a news media. Java et al. (2007) suggested a taxonomy of intentions of Twitter user microblogging posts: daily chatter, conversation, information sharing, news reporting. This indicates that communication plays a crucial role in Twitter usage. This implies that Tweets do not necessarily need to be influenced by or to be related to the location where the Twitter users post their messages. This needs to be considered by any geographical analysis approach using Twitter data.

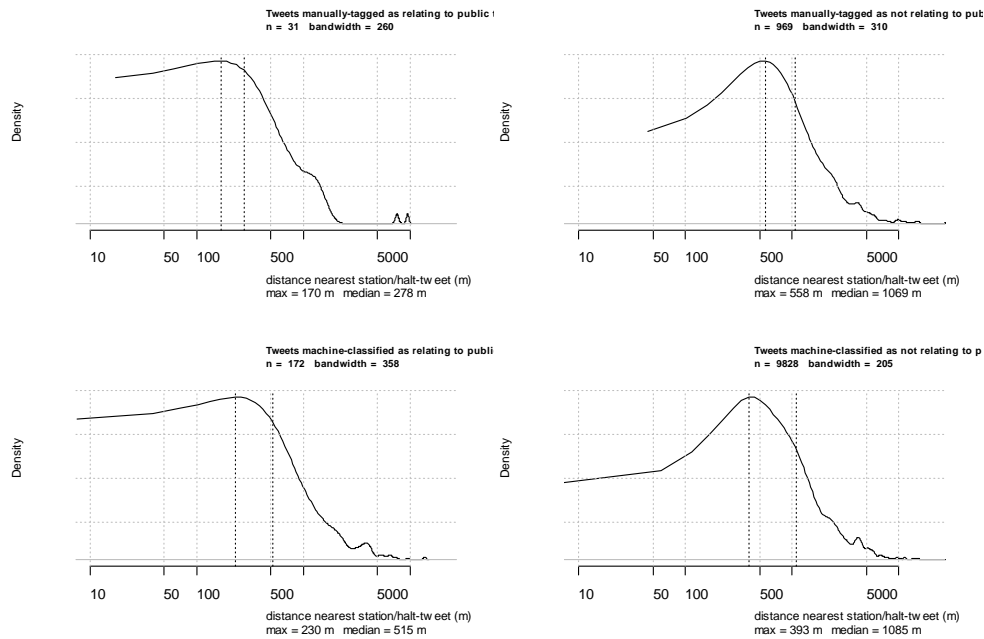


Figure 4. Density plots for the distribution of Twitter microblogging posts that related to the topic 'public transport' with regard to their distance to the nearest train station on logarithmic scale. Left: messages that are related to public transport, right: messages that are not related to public transport, top: results of manual classification, bottom: results of machine classification.

In order to estimate how much mobile microblogging contents are influenced by nearby points of interest (POI), we use methods of natural language processing. As a first example we have developed and trained a Maximum Entropy classifier that distinguishes Tweets that are (not) related to public transport. This enables us to analyse the correlation between the location where mobile microblogging posts of this topic have been generated and their closest train station, which may be assumed to be a strong factor that triggers posts about the topic 'public transport'. Figure 4 shows a density plot of the according probability distribution for both the manually classified and the machine classified microblogging posts. It can be seen that there is a difference between the distributions of distances of Tweets that are (not) related to the topic of 'public transport'. Furthermore, it can be seen that the difference is also visible for the machine classification. In general, mobile microblogging posts that are related to the topic 'public transport' tend to be nearer to the closest train station than posts that are not related to this topic. However, the effect is not as strong as it may be assumed. The reason for that may be that Twitter, to a certain extent, is a social network, whose contents is influenced by many factors of which geo-spatial context is only one.

References

- Hahmann, S. & Burghardt, D. (2011): Maple – a Web Map Service for Verbal Visualisation Using Tag Clouds Generated from Map Feature Frequencies. In: *Advances in Cartography and GIScience*. Selection from ICC 2011, Paris, Springer, 3-12.
- Hollenstein, L. & Purves, R.S., (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. In: *Journal of Spatial Information Science*, 1 (1), 21–48.
- Java, A., Finin, T., Song, X. & B. Tseng, (2007). Why We Twitter: Understanding Microblogging Usage and Communities. In: *Proceedings of the 9th WebKDD and 1st SNA-KDD workshop on Web mining and social network analysis*.
- Jones, C.B., Purves R.S., Clough P.D. & H. Joho (2008): Modelling vague places with knowledge from the Web, *International Journal of Geographical Information Science*, 22 (10), 1045–1065.
- Kwak, H., Lee, C., Park, H. & S. Moon (2010): What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on WWW*, 591-600.